

What Chinese readers see in character components

James Myers, National Chung Cheng University (rejected conference abstract)

Intuitions about the phonotactic well-formedness of novel words depend on structured knowledge about their formal similarity with real words (Albright, 2009). To explore whether orthographic symbols are also parsed in a structured way, we analyzed acceptability judgments for single-component nonlexical Chinese characters in terms of lexical neighbors and universal visual/manual constraints.

Test items were selected from the extinct Khitan script (Kane, 2009) as set in a Song typeface (<http://www.babelstone.co.uk>); see examples in Table 1. In Experiment 1, 23 native Chinese-speaking traditional character readers were presented with 136 target items and asked to write down the most similar-seeming real character and its degree of similarity with the target. The structured nature of these responses can be seen by comparing them with the characters identified by Acrobat's optical character recognition (OCR), which was sensitive solely to the distribution of black and white, ignoring reader-salient stroke groups, as seen in *a-d*.

Experiment 2 presented the same test items to a new set of 36 native Chinese-speaking traditional character readers, who gave speeded yes/no acceptability judgments. Statistical analyses showed that acceptability (shown as cross-reader acceptance rate in Table 1) was higher if more readers in Experiment 1 had agreed on the closest lexical neighbor and had given it a higher similarity score, but having a close match still did not guarantee high acceptability, as in *e, g, h*. Acceptability was also significantly higher the more strokes in the item; traditional Chinese readers seem to expect real characters to be visually complex. There was also a trend consistent with a visually motivated universal disfavoring cardinal/oblique axis combinations (Morin, 2018), as in *f* and *i*, though this did not reach statistical significance. Acceptability also tended to conform to a manually motivated universal (van Sommers, 1984): items with more strokes starting from another stroke, as in *d*, were judged more acceptable than those with more strokes ending at another stroke, as in *h*, though this effect was statistically significant only when neighbor-related predictors were excluded from the analysis.

Together these results reveal rich structured knowledge about what makes even simple Chinese character components well- or ill-formed, consistent with a view of writing systems as more similar to natural language than linguists have traditionally acknowledged (Meletis, 2018).

Table 1. Sample test items and results of Experiments 1 and 2.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
Item	美	冊	茁	彡	木	𠂇	女	止	友
Neighbor	美	冊	由	万	木	牟	女	止	及
OCR	芙	冊	茁	气	木	牟	玄	止	友
Acceptability	.92	.80	.64	.47	.39	.26	.19	.11	.06

References

- Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26(1), 9-41.
- Kane, D. (2009). *The Kitan language and script*. Leiden/Boston: Brill.
- Meletis, D. (2018). What is natural in writing? Prolegomena to a *Natural Grapholinguistics*. *Written Language & Literacy*, 21(1), 52-88.
- Morin, O. (2018). Spontaneous emergence of legibility in writing systems: The case of orientation anisotropy. *Cognitive Science*, 42(2), 664-677.
- van Sommers, P. (1984). *Drawing and cognition: Descriptive and experimental studies of graphic production processes*. Cambridge: Cambridge University Press.